Otimização não-linear irrestrita: um estudo teórico e computacional

Unrestricted nonlinear optimization: a theoretical and computational study

Felipe Romero da Mata (orientado)*, Tatiane Cazarin da Silva (orientadora)†

RESUMO

O presente trabalho tem como objetivo realizar um estudo teórico e computacional acerca dos métodos de otimização irrestrita aplicados a minimização de funções. Dessa forma, pode-se investigar qual é a solução mais eficiente, computacionalmente, para a solução de um determinado problema. Para isso, foram desenvolvidos códigos computacionais, implementados no software Matlab, que se diferenciavam quanto ao método de otimização, tais como Cauchy, Newton, Quase-Newton e Gradientes Conjugados, assim como à escolha do tamanho de passo. Os métodos teóricos foram aplicados a uma classe de funções, sendo escolhida uma delas para a discussão computacional e numérica. Os dados alcançados no estudo apontaram que a eficiência do método varia de acordo com cada tipo de problema e cabe ao pesquisador verificar qual método adotar em cada caso.

Palavras-chave: Cauchy, Newton, gradiente

ABSTRACT

The present work aims to carry out a theoretical and computational study about the unrestricted optimization methods applied to function minimization. Thus, it is possible to investigate which is the most computationally efficient solution to solve a given problem. For this, computational codes were developed, implemented in Matlab software, which differed in terms of the optimization method, such as Cauchy, Newton, Quase-Newton and Conjugated Gradients, as well as the choice of step size. The theoretical methods were applied to a class of functions, one being chosen for the computational and numerical discussion. The data obtained in the study showed that the efficiency of the method varies according to each type of problem and it is up to the researcher to verify which method to adopt in each case.

Keywords: Cauchy, Newton, gradient

1 INTRODUÇÃO

Otimizar consiste em buscar a solução mais adequada para um problema, dentre diversas outras soluções alternativas. O critério de avaliação recebe o nome de função objetivo, é esta que se busca otimizar, ou seja, maximizar ou minimizar. Em alguns casos, as soluções alternativas podem estar sujeitas a supressão, o que indicaria a presença de restrições que devem ser respeitadas (GOLDBARG e LUNA, 2000).

Pode-se usar a programação não linear quando se tem a função objetivo e/ou restrições não-lineares. Os problemas de otimização podem ser aplicados em diversas áreas como engenharia, administração, logística, transporte, economia, biologia ou em qualquer outra área onde seja possível construir modelos matemáticos representantes dos sistemas em análise (AVRIEL, 1976).

* Engenharia Civil, Universidade Tecnológica Federal do Paraná, Campo Mourão, Paraná, Brasil; feliperomero@alunos.utfpr.edu.br
† Universidade Tecnológica Federal do Paraná, Campus Campo Mourão; tatianecazarin@utfpr.edu.br

Ao longo deste artigo serão explanados métodos clássicos de otimização irrestrita, como: Cauchy, Newton, Quase-Newton e Gradientes Conjugados, e também métodos exatos e inexatos de cálculo de passo, como: Seção Áurea, Busca linear exata e Condição de Armijo e, a partir disso, eles serão combinados e comparados quanto a sua eficiência. Tais métodos são desenvolvidos por meio de um processo iterativo. Considera-se um ponto inicial x_0, e obtém-se um ponto melhor x_1, ou seja, em um problema de minimização, um ponto que diminua o valor da função objetivo. Esse processo é repetido gerando uma sequência na qual a função objetivo decresce (FERRAZ, 2017). Em suma, o objetivo deste trabalho é fazer um estudo teórico e computacional a respeito dos problemas de otimização irrestrita, explicitando os métodos que são destacados na literatura.

2 MÉTODO

Um problema de otimização irrestrito pode ser escrito como

$$\begin{aligned} &minimizar\ f(x) \\ &sujeito \text{ à } x \in R^n \end{aligned} \quad (1)$$

em que a função $f : R^n \to R$, denominada função objetivo, é de classe C^n. Quando se considera tal problema, o objetivo é determinar minimizadores locais de f.

Caso a função objetivo esteja definida sobre um subconjunto Ω pertencente ao R^n, sujeito a restrições de igualdade e/ou desigualdade, haverá então um caso de otimização restrita. Para tais formulações, são aplicados métodos de penalização, que o transformam num problema irrestrito, como: pontos interiores, multiplicadores de Lagrange, Programação Quadrática Sequencial.

Em alguns casos de otimização pode ser necessário encontrar o máximo de uma função, como por exemplo quando se quer maximizar o lucro de um investimento. Porém, da perspectiva matemática, não há diferença entre buscar a minimização ou a maximização de uma função, uma vez que maximizar f é equivalente a minimizar $-f$. Sendo assim, neste artigo abordaremos a otimização aplicada a minimização de funções.

Um algoritmo básico de otimização irrestrita determinará, a partir de cada ponto, uma direção para dar o próximo passo. Como o objetivo é minimizar f, faz sentido que a função decresça na direção escolhida. Um bom algoritmo será aquele onde a função convergirá para um mínimo no menor tempo possível.

As direções que formam ângulos superiores a 90º com o gradiente de f avaliado no ponto $\bar{x}$, são direções de descida para f a partir de tal ponto. O que diferencia os métodos de otimização é a escolha da direção de descida d e a forma com que é calculado o tamanho do passo t a ser dado. Após escolher a direção segue-se para o próximo passo, definir o quanto caminhar nesta direção. É possível dar o passo completo, ou seja, considerar o tamanho $t = 1$, porém esta escolha pode não ser a melhor. Sendo assim, existem os métodos denominados métodos de busca que podem ser empregados para este fim.

Tendo isso em vista, agora será apresentado um modelo de algoritmo básico genérico (Algoritmo 1) que minimiza uma função $f : R^n \to R$, disponível em Ribeiro e Karas (2011) e na sequência serão abordados os tipos de cálculo para se encontrar o tamanho de passo t e os métodos que determinam a direção d.

Algoritmo 1: Algoritmo genérico:

Dados de entrada: $x_0 \in R^n$
$k = 0$
Repita enquanto $\nabla f(x_k) \neq 0$
Defina d_k *tal que* $\nabla f(x_k)^T d < 0$
Obtenha $t_k > 0$ *que minimiza* $f(x_k + t_k d_k)$
Faça $x_{k+1} = x_k + t_k d_k$

$k = k + 1$
Retorna x_k, k

A escolha tanto da direção d_k quanto do tamanho do passo t_k, na iteração k, tem que ser feita de modo a garantir a convergência de f. Os métodos abordados neste trabalho que dão o tamanho do passo podem ser exatos, como: Busca Linear Exata e Seção Áurea e também inexatos, como: Condição de Armijo. Enquanto os métodos utilizados para se encontrar a direção serão os métodos de Cauchy, Newton, Quase-Newton e Gradientes-Conjugados. A seguir tais métodos são apresentados.

2.1 Escolha de direção

2.1.1 Método de Cauchy

O método de Cauchy, também chamado de Método do Gradiente ou ainda Método da Descida Máxima, é um processo iterativo onde cada etapa avança na direção oposta ao vetor gradiente da função objetivo no ponto vigente, ou seja, no Algoritmo 1 tem-se que:

$$d_k = -\nabla f(x_k) \tag{2}$$

O que justificaria tal escolha é o fato do vetor gradiente avaliado em um ponto x_k apontar para a direção com maior crescimento de f. Dessa forma, ao andar para a direção oposta, espera-se que o vetor forneça a direção de maior descida da função objetivo (FERRAZ, 2017).

2.1.2 Método de Newton

A ideia do método de Newton é utilizar métodos iterativos para aproximar uma função F por seu polinômio de Taylor. O Algoritmo de Newton pode não estar bem definido, caso a matriz Hessiana $\nabla^2 f(x_k)$ seja singular. Além do mais, mesmo que o passo d_k seja calculado, essa direção pode não ser de descida. Contudo se $\nabla^2 f(x_k)$ é definida positiva, então o passo d_k está bem definido e é uma direção de descida (FERRAZ, 2017), dado por:

$$d_k = -\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k). \tag{3}$$

2.1.3 Método de Quase-Newton

Métodos quase-Newton, assim como o método de máxima de descida, necessitam que apenas o gradiente da função objetivo esteja disponível em cada iteração. Ao medir as mudanças no gradiente de uma iteração para outra, eles tentam construir um modelo para a função objetivo bom o bastante para produzir convergência.

De acordo com Izmailov e Solodov (2007), a direção utilizada pelo método de Quase-Newton para minimizar uma função f é dada por:

$$d_k = H_k \nabla f(x_k) \tag{4}$$

onde $H_k \in R^{n \times n}$ é definida positiva.

2.1.4 Método dos Gradientes-Conjugados

A ideia deste método consiste em, dado um ponto inicial x_0, adota-se uma primeira direção que é a inversa do gradiente e a partir desta, todas as outras direções são A-conjugadas entre si. Duas direções d^i e d^j, são ditas direções A-conjugadas se $\left(d^i\right)^T A d^j = 0$, para todos $i, j = 0, 1, \dots, k$, com $i \neq j$. O Algoritmo começa definindo $d_0 = -\nabla \mathrm{f}(x_0)$, e a partir da primeira iteração constrói uma sequência de direções dadas por

$$d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d^k \tag{5}$$

em que o valor de β_k, conforme Pamplona (2019), podem ser definidos por:

$$\beta_k = \frac{\nabla f(x_{k+1})^T \cdot (\nabla f(x_{k+1}) - \nabla f(x_k))}{\nabla f(x_k)^T \cdot \nabla f(x_k)} \text{ ou } \beta_k = \frac{\nabla f(x_{k+1})^T \cdot (\nabla f(x_{k+1}))}{\nabla f(x_k)^T \cdot \nabla f(x_k)}, \tag{6}$$

O que permite que as direções geradas sejam A-conjugadas.

2.2 Escolha do tamanho do passo

2.2.1 Busca linear exata

Com o objetivo de determinar o tamanho do passo, t_k utilizado no Algoritmo 1, a busca linear exata parte de um ponto x^k em uma direção já definida por um dos métodos anteriormente citados. O tamanho de passo t_k é encontrado pela minimização da função $\varphi(t_k) = f(x_k + t_k d_k)$. Fazendo $\nabla\varphi(t_k) = 0$, obtêm-se a fórmula fechada para o passo, que no caso de funções quadráticas $f\colon \mathbb{R}^n \to \mathbb{R}$, definidas por $f(x) = \frac{1}{2}x^T A x + b^T x + c$, tal passo é dado por:

$$t_k = \frac{\nabla f(x_k)^T \cdot d}{d^T \cdot A \cdot d} \tag{7}$$

2.2.2 Seção Áurea

Este método, de acordo com Ferraz (2017), encontra um ponto próximo de um minimizador para uma função φ, com uma precisão ε desejada. Seu algoritmo pode ser dividido em duas fases. Na primeira obtém-se um intervalo $[a, b]$ que contém um minimizador de φ. A ideia desta etapa é considerar um intervalo inicial $[0,2\rho]$, com $\rho > 0$, e ampliá-lo, deslocando para a direita, até que um crescimento de φ seja detectado. Na segunda fase, o intervalo $[a, b]$ é reduzido, através do descarte de subintervalos, até que reste um intervalo de tamanho suficiente para que uma precisão ε seja alcançada.

2.2.3 Condição de Armijo

A condição ou busca de Armijo, de acordo com Junior et al (2013), consiste em obter uma redução da função objetivo ao longo da direção adotada sem tentar minimizá-la. A condição de Armijo assegura um decréscimo monótono dos valores da função objetivo ao longo do processo iterativo.

Considerando então um ponto $\bar{x} \in R^n$,uma direção de descida $d \in R^n$ e $\sigma \in (0,1)$, basicamente a busca de Armijo encontra $\bar{t} > 0$ tal que

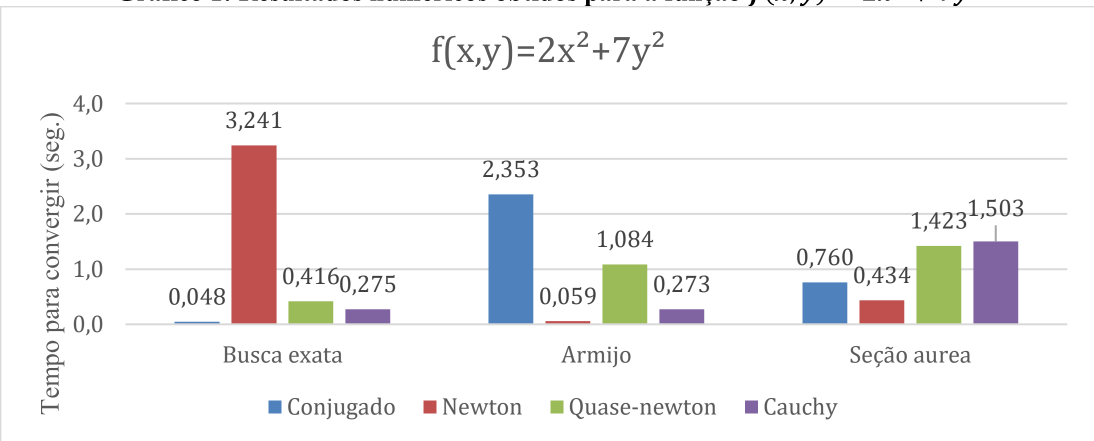
$$f(\bar{x} + \bar{t}d) \leq f(\bar{x}) + \sigma\bar{t}\nabla f(\bar{x})^T d. \tag{8}$$

3 RESULTADOS

Até agora foram expostos métodos de otimização e de cálculo do tamanho do passo, mas falta esclarecer um ponto: quando devo utilizar cada um? Para responder esta questão foi escolhida uma função para ser otimizada, variando apenas o método empregado, de modo que se possa comparar a velocidade de convergência em cada caso. A função escolhida para os testes de velocidade de convergência por métodos irrestritos foi $f(x,y) = 2x^2 + 7y^2$.

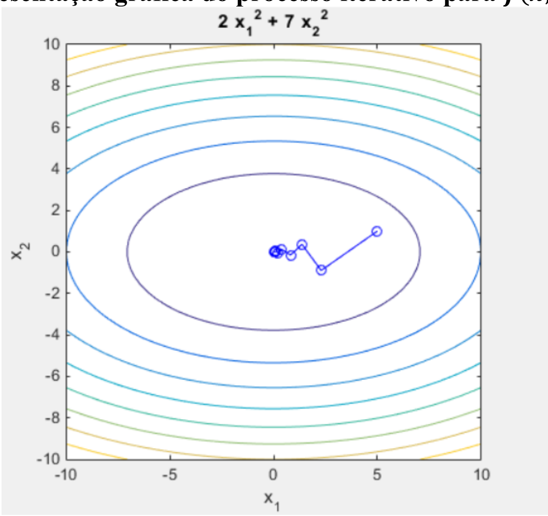
A implementação dos métodos foi realizada no software Matlab. Como parâmetros foram utilizados: Ponto inicial (5, 1) e como critério de parada $\nabla f(x,y) \leq 10^{-4}$. Os resultados obtidos através da aplicação dos métodos considerados, podem ser vistos no Gráfico 1 a seguir.

Gráfico 1: Resultados numéricos obtidos para a função $f(x,y) = 2x^2 + 7y^2$



Fonte: Autoria própria (2021)

Gráfico 2: Representação gráfica do processo iterativo para $f(x,y) = 2x^2 + 7y^2$



Fonte: Autoria própria (2021)

Com a análise dos dados obtidos pode-se notar que o Método de Newton combinado com a Busca Linear Exata foi o que levou mais tempo para alcançar a convergência, levando aproximadamente 3,24 segundos.

Porém alterando o método da escolha de passo para Condição de Armijo e Seção Áurea este tempo é reduzido consideravelmente, sendo Armijo o mais eficiente dos três, convergindo em 0,06 segundos.

Ao observar o método Quase-Newton nota-se que a mudança da escolha de passo, para esta função específica, não foi tão relevante como no caso anterior. A função convergiu mais rapidamente utilizando o método da Busca Linear exata, levando 0,42 segundos para convergir. E o caso mais lento foi quando se utilizou a Seção Áurea, neste caso foram necessários 1,42 segundos.

No caso do Gradiente-Conjugado, a variação mais relevante foi quando se utilizou a Condição de Armijo, neste caso foram 2,35 segundos para convergir. Quando se utiliza a Seção Áurea e a Busca Exata este tempo diminuiu, sendo 0,76 e 0,05 segundos para cada um, respectivamente, sendo este segundo a combinação que alcançou a convergência mais rápida em todos os testes realizados.

E por fim, ao utilizar o Método de Cauchy combinado com Busca Exata e Condição de Armijo o tempo de convergência foi de 0,27 segundos para ambos os casos, porém ao utilizar a Seção Áurea ocorreu um aumento no tempo de convergência, levando 1,50 segundos para convergir.

De maneira geral, dada a natureza do problema a ser minimizado, os métodos podem variar em eficiência, sendo assim, eles precisam ser adaptados e reavaliados em cada análise para que se possa escolher o mais apropriado para o momento.

4 CONCLUSÃO

Através das pesquisas desenvolvidas e descritas ao longo deste trabalho, foi possível identificar a importância dos estudos de otimização em diversas áreas do conhecimento, sobretudo, na Engenharia Civil. Sempre que se busca fazer o melhor com a menor quantidade de recursos possíveis podemos recorrer aos algoritmos de otimização, uma vez que eles podem fornecer os melhores valores para alcançar tal resultado.

Este estudo contemplou apenas alguns métodos possíveis para se otimizar uma função, cada um com suas vantagens e desvantagens. Dessa forma fica a cargo de profissionais especializados escolher o método de otimização mais adequado para cada situação. Essa escolha pode ser feita através de testes de comparação, como feito neste trabalho, pois ele revela de qual método se chegou mais rapidamente ao resultado desejado.

REFERÊNCIAS

AVRIEL, Mordecai. **Nonlinear programming:** Analysis and Methods. New Jersey: Prentice-Hall, 1976.

FERRAZ, Bruna Alves. **Métodos computacionais de otimização**. Instituto de Geociências e Ciências Exatas, UNESP - Rio Claro, 2017.

GOLDBARG, Marco Cesar; LUNA, Henrique Pacca Loureiro. **Otimização Combinatória e Programação Linear**: Modelos e Algoritmos, Editora Campus, Rio de Janeiro, 2000.

IZMAILOV, Alexey.; SOLODOV, Mikhail. **Otimização:** métodos computacionais. Volume 2. Rio de Janeiro: IMPA, 2007.

JUNIOR, Marlon Luiz Dal Pasquale; SANTOS, Solange Regina; PERIÇARO, Gislaine Aparecida. **Métodos Newton e Quase-Newton para otimização irrestrita**. Campo Mourão: VIII EPCT, Unespar/Fecilcam, 2013.

RIBEIRO, Ademir Alves; KARAS, Elizabeth Wegner. **Um curso de otimização**. Curitiba, 2011.